Coverage of patient safety terms in the UMLS Metathesaurus

Aziz A. Boxwala, MBBS, PhD[1,2], Qing T. Zeng, PhD[2], Anthony Chamberas, MS[1],
Luke Sato, MD[1], Meghan Dierks, MD[3,4]

[1]Risk Management Foundation of the Harvard Medical Institutions, Cambridge, MA
[2]Decision Systems Group, Brigham & Women's Hospital, Harvard Medical School, Boston, MA
[3]Center for Clinical Computing, Beth Israel Deaconess Medical Center, Harvard Medical School,
Boston, MA, [4]Clinical Decision Making Group, MIT, Cambridge, MA

The integration and large-scale analyses of medical error databases would be greatly facilitated by the use of a standard terminology. We investigated the availability in the UMLS metathesaurus of concepts that are required for coding patient safety data. Terms from three proprietary patient safety terminologies were mapped to the concepts in UMLS by an automated mapping program developed by us. From these candidate mappings, the concept that matched its corresponding term was selected manually. The reliability of the mapping procedure was verified by manually searching for terms in the UMLS Knowledge Source Server. Matching concepts in UMLS were identified for less than 27% of the terms in the study dataset. The matching rates of terms that describe the type of error and the causes of errors were even lower. The lack of such terms in the existing standard terminologies underscores the need for development of a standard patient safety terminology.

INTRODUCTION

According to a report published by the Institute of Medicine in 1999, more than one million preventable adverse events occur nationwide resulting in tens of thousands of deaths each year [1]. In order to better understand the frequency, types, and causes of medical errors that occur during the management of patients, healthcare institutions have deployed systems for reporting incidents [2, 3]. These systems enable staff to report incidents that caused or had the potential to cause harm to the patient. Other types of reports such as those from formal investigations of incidents [4] and malpractice claims [5] provide further information on the nature of medical errors.

To avoid bias and assure the most faithful variable selection, data should be aggregated and analyzed from as many different medical disciplines and institutions as possible. To achieve this, researchers and policymakers have advocated the creation of statewide or nationwide databases of error and near miss reports [6].

A common terminology, which is required for encoding the reports in a shared, large-scale database, does not exist. Preliminary reports indicate that existing controlled clinical terminologies such as SNOMED, ICD-9 CM, or CPT [7-9] do not contain terms relating to medical errors and their attributes [10]. Developers of incident reporting systems have created proprietary and application-specific terminologies for use in their systems, but there is tremendous heterogeneity across these sources, and a common reference model does not exist.

As an initial step in developing a standard reference model for patient safety terminology, we performed a comprehensive audit of standard clinical terminologies contained within the Unified Medical Language System (UMLS) metathesaurus to determine the extent to which existing terms for patient safety applications (incident reporting systems, insurance industry risk codes, etc.) are covered. We selected three representative patient-safety-related terminologies, and mapped terms to concepts in the UMLS metathesaurus [11]. The mapping was performed using a software tool that generated candidate concepts that matched a term. An Informatician selected the correct concept for a term from the candidate concepts.

METHODS AND MATERIALS

Terminologies used

For this study, we analyzed three different patient safety terminologies that were among the most comprehensive and representative of the patient safety issues. The sources and their general attributes are summarized in Table 1. DoctorQuality Inc.'s Risk Prevention and Management System™ (RPM) uses a broad but proprietary terminology for encoding incident reports from a variety of clinical domains. The Risk Management Foundation (RMF), a malpractice insurer, has developed and uses a proprietary terminology for encoding medico-legal claims data. The NCC-MERP taxonomy is used for coding medication-related errors [12] and is representative of terminologies for a specific application domains. Terms in the latter group of terminologies are fine-grained and narrowly focused on their respective domains. All three terminologies organize terms into categories, and terms within a category are represented largely in *is-a* hierarchies. The terminologies have not been developed using formal knowledge representation approaches, however, so there is some inconsistency in the relationships between major and minor terms within a category.

Table 1. A partial listing of patient-safety-related terminologies with categories, sample terms, and the number of terms from the category used in the study dataset

Selected categories	Sample terms	No. of terms
NCC-MERP's Taxonomy of Medication Errors		273
Setting	Adult day health care	56
Product (Drug) information	Tablet	20
Personnel involved	Licensed Practical Nurse	23
Type (of error)	Dose omission	29
Causes	Written miscommunication	98
Contributing factors	Lack of availability of health care professional	21
DoctorQuality Inc.'s Risk Prevention and Management System's Terminology		518
Adverse clinical event	Fall	213
Administrative incident	Chart unavailable	100
Contributing factors	Distractions in the environment	61
Level of impact	Near death event	10
Medication type	Antidepressant	54
Roles	Respiratory therapist	37
Risk Management Foundation's Malpractice Claims Codes		471
Allegations	Inappropriate transfer	91
Location	Radiation therapy	64
Services	Radiology	61
Employee	Chiropractor	53
Risk management issues	Lack of any consent	149

For the purpose of this study, we considered the content of the three terminologies (two broad and one domain-specific) as sufficient and representative. Inspection of other broad terminologies indicated significant overlap with the selected terminologies (RMF and RPM). We chose the NCC-MERP taxonomy as representative of domain-specific terminologies. Inclusion of other domain-specific terminologies (e.g., MERS-TM [2] for transfusion-related incidents) would have added more domain-specific terms to the dataset but we believed would be unlikely to change the nature of the conclusions. We also did not consider taxonomies that to our knowledge are not being used for coding error reports.

Preparation of dataset

Consolidation of these three terminologies produced an initial study dataset of 1262 terms. The study dataset was refined in several ways. First, free-standing, but non-informative terms such as *Other*, *Unknown*, and *Not applicable* were removed. Terms that were narrower than a parent term only by virtue of non-specific modifiers such as *Other* were also removed from the dataset (e.g., *Equipment, other)*. Codes for field names (e.g., "The Event" in the NCC-MERP taxonomy) and terms referring to names of fields for free-text (e.g., name of manufacturer) and ordinal data (e.g., dose) also were removed. After this filtering, a total of 1140 terms remained in the dataset.

In several cases, the full meaning of a term was not correctly represented without considering the context of the term, i.e., its ancestors. For example, in the term hierarchy below, the term *Community* is intended to mean community pharmacy.

24.13 Pharmacy
24.13.1 Community

All terms in the dataset were inspected; for 244 terms, a composite term was constructed by concatenating a term with its ancestor terms in order to convey its full meaning. For example, the composite term for the example above is *Community Pharmacy*.

Finally, related term categories were consolidated into a smaller number of new categories. The mapping between the categories is shown in Table 2.

Mapping to UMLS

We used the 2003 AA edition as the UMLS metathesaurus source. Each concept in the metathesaurus has a unique identifier that is associated with one or more concept names. The 2003AA edition contains 875,255 concepts and 1,773,525 English language concept names derived from over 100 different source terminologies. Each term in the metathesaurus is assigned to one or more semantic types from the UMLS semantic network. The metathesaurus was loaded in a MySQL relational database. Indexes were built on concept name fields to improve the efficiency of queries.

In order to reduce the burden of manually mapping 1140 terms in the dataset to UMLS concepts, a program was written in Java to automatically find concepts that potentially mapped to each term. These 'candidate mappings' generated by the program then

were reviewed manually to confirm that the program's mapping was correct

Automatic mapping

The mapping algorithm is based on the Minimal Representable Units Method (MRUM) developed by Zeng et al [13], but includes a few refinements. The algorithm first attempts to perform a high-specificity *exact match* of an input term to a metathesaurus concept name. This phase of the algorithm is performed in the following sequence of steps:

1. The input term is normalized using the lexical tools (specifically, the *norm* API) provided with UMLS. The default parameters of the norm API were used. Normalization includes such steps as removal of stop words (e.g., of, the), removing inflections, changing term to all lower case, and alphabetical ordering of words in the term.

2. The normalized term is matched with concept name entries from the *mrxns.eng* file in the metathesaurus. This file contains normalized forms of concept names in the English language only. An input term may match more than one concept due to the presence of homonyms in UMLS. The unique set of matched metathesaurus concepts is selected.

3. The semantic type of each selected concept is compared to the category of input term. If one of the semantic types of the concept matches the input term's category, that concept is saved in the results file. The matching is based on a mapping of input term categories to the semantic types in UMLS that we created (column 3 of Table 2). This step eliminates any matches that are not semantically compatible.

If, after step 3, an input term does not map to any concept in the UMLS metathesaurus, the algorithm performs a *partial match*. This match procedure is designed to identify potential matches to the complete term and matches to parts of the term using a high-sensitivity, low-specificity matching method. The partial match is performed as follows:

4. Partial strings are generated from the input term by selecting all combinations of sequential words in the term. Thus, from the term *failure to diagnose,* the following strings are generated: *failure*, *failure to*, *failure to diagnose*, *to*, *to diagnose*, and *diagnose*.

5. Each string is normalized as in step 1. The normalized string is matched to concepts as in step 2. Since these are partial strings, we do not match the category of the whole input term to the semantic type of the matched concept.

6. The mapped concepts for all the partial strings of a term are combined and duplicate concepts are removed. This step is necessary since normalization of the partial strings can result in duplicate strings (e.g., *failure* and *failure to* both normalize to *failure*). The set of unique matched concepts is stored in the results file.

Manual selection and verification

Results of the automated mapping algorithm were loaded in an Access database. Forms were designed to display the matches through a user interface. Using this interface, the correctly matched concept for each term was selected manually by an Informatician-Physician expert. Further, for each term in the dataset, the algorithm-driven match was marked as 'correct' (C) if the input term mapped correctly and completely to one of the candidate UMLS concepts. The match was judged 'partially correct' (PC) if only part of the input term mapped correctly to a candidate

Table 2. Aggregation of original term categories into a smaller number of categories. The column labeled STY lists the identifiers of the UMLS semantic types that match the new category. The last two columns on the right show the number of terms, the average number of words and the standard deviation of the normalized terms in the new category.

New category	Old categories	STY	Count	Num words (std. dev.)
DRG	Product (Drug) information	All	20	2.25 (1.45)
ERR	Allegations, Administrative incident, Adverse clinical event, Type of error	All	433	2.58 (1.23)
FAC	Contributing factors, Causes, Risk management issues	All	329	4.30 (2.31)
LOC	Setting, Location	T073, T090, T091, T092, T093	120	2.12 (0.97)
MED	Medication type	T103-T121, T124-T130, T195-T197, T200	54	1.50 (0.72)
OUT	Level of impact	All	10	3.10 (1.29)
ROL	Personnel involved, Roles, Employee	T096, T097, T098, T099, T100, T101	113	1.71 (0.75)
SVC	Services	T090, T091, T092, T093	61	1.52 (0.62)

concept. The match was deemed incorrect (IC) when the input term did not match any of the candidate concepts. For some terms, the program did not generate candidate concepts (NM).

For term in the NM, PC, or IC category, a second Informatician-Physician expert manually searched the UMLS Knowledge Source Server (KSS) to verify that a correct concept did not exist in UMLS. This manual expert search was performed using the normalized string index, normalized word index, and word index parameters of KSS.

Analysis of data

A binomial logistic regression test was performed on the data using the SPSS program. The output variables of this statistical test were (1) the *type of automatic matching algorithm* (exact or partial); and (2) the *correctness of the mapping* (correct or not correct; the latter groups PC, IC, and NM). The input variables were (1) the categories of the terms; (2) the source terminology; and (3) the number of words in the term. A previous study had demonstrated the number of words in a text strings to be a significant factor in mapping between text strings and UMLS concepts [14]. In our analysis, we used the number of words in the normalized term since our program matched normalized terms.

RESULTS

The program mapped 243 terms (21.3%) to 274 UMLS concepts (average of 1.13 concepts per term) using the exact matching algorithm, mapped 858 terms (75.3%) to 4761 concepts (average of 5.55 concepts per term) using the partial matching algorithm, and could not map (NM) 39 (3.4%) terms using either approach.

Table 3. Performance of the automatic mapping tool. Note that the program did not match 39 terms to any concepts. Keys: C=Correct, PC=Partially Correct, IC=Incorrect.

Automatic match	Correctness			
	C	PC	IC	Total
Exact	230	10	3	243
Partial	74	721	63	858
Total	304	731	66	1101

Table 3 shows the results of the automatic mapping program and the manual verification. Of note are that (1) the exact matching algorithm performed correctly for 230 of the 243 (95%) terms for which it found matches; and (2) 304 concepts (26.7%) were correctly found by the automatic matching procedure through exact and partial matching approaches.

The number of words in the term had an impact on the *correctness of the mapping*. No correct matches

to a UMLS concept were found for terms containing greater than 7 words. These terms were excluded from further analysis. The probability of finding a correctly matched concept decreased as the number of words increased (p<0.001). The source terminology did not have an impact on the probability of finding a correctly matched concept. The categories ERR (p<0.001) and FAC (p<0.001) decreased the probability of finding a correctly matched concept and the category MED (p=0.02) increased the probability of finding a correctly matched concept as compared to the other categories. Table 4 shows the probability of correct mapping for different categories and lengths. Similar results were obtained for the *type of automatic match* outcome. The notable difference was that there never was an exact match for any term that had more than four words.

Table 4. Probability of finding correctly matched concepts for terms from different categories given the number of words in the term.

Categ.	Number of words in the term						
	1	2	3	4	5	6	7
ERR	0.46	0.27	0.13	0.06	0.03	0.01	0.01
FAC	0.45	0.26	0.13	0.05	0.03	0.01	0.01
LOC	0.63	0.41	0.23	0.11	0.05	0.02	0.01
MED	0.88	0.76	0.58	0.36	0.20	0.09	0.04
Others (base)	0.75	0.56	0.35	0.19	0.09	0.04	0.02

Table 5 shows the number and percentage of terms by category for which a correct matching concept was found. For 16% of ERR terms and 9% of FAC terms, matching concepts were found. Terms from MED matched at a rate of 80%.

Table 5. Frequency distribution by categories of the terms that were matched to UMLS concepts.

Category	Total terms	Correct (C)	Not correct (IC+PC+NM)
DRG	20	10 (50%)	10
ERR	433	68 (16%)	365
FAC	329	28 (9%)	301
LOC	120	44 (37%)	76
MED	54	44 (80%)	10
OUT	10	1 (10%)	9
ROL	113	67 (59%)	46
SVC	61	42 (69%)	19

Manual search in the KSS was performed for 39 terms that were not mapped by the program (NM), and 50 terms randomly selected from 798 partially matched (PC, n=721) or incorrectly matched terms (IC, n=63). Of the 39 NM terms, the manual search found concepts for 12 terms (31%). Of the 50 PC and IC terms, we found correct concepts for 8 terms (16%). For the combined set of 89 terms, correct con-

concepts were found for 6 of 39 (15%) ERR terms and none of the 23 FAC terms.

DISCUSSION AND CONCLUSIONS

Based on these results, we determined that there are no matching UMLS concepts for more than 70% of the patient safety-related terms from the dataset. Not surprisingly, matched terms were more likely to be from the clinical categories ROL, SVC and MED rather than the ERR, FAC and OUT categories, reflecting the constituency of the source terminologies of the UMLS metathesaurus. This study reinforces preliminary findings by others [10] of the inadequacy of existing standard terminologies for coding patient safety data and underscores the need for the adoption or development of a standardized patient safety terminology. However, adoption of an existing terminology may not be successful because no single terminology appears comprehensive enough to cover the range of clinical domains and applications. Furthermore, although not the primary focus of this study, existing terminologies appear inconsistent and highly variable in terms of granularity. This analysis shows that any common terminology development effort should focus on terms for error type (ERR) and the causative factors (FAC).

We believe that the low percentage of terms for which additional concepts were found by manual search demonstrated the reliability of the mapping procedure. The additional matches will not alter the conclusions about the lack of matching concepts for terms from the ERR and FAC categories.

Using the automatic mapping program to screen candidate matching concepts for more than 1000 terms reduced what otherwise would have been a significant manual effort. The two-step approach of performing a high-specificity exact match (which largely produced correct matches for manual selection) followed by a high-sensitivity partial match (which narrowed choices for the remainder of the terms) also reduced the labor involved in this study.

The number of words in the term, which is indicative of the term's complexity, significantly affected the ability to match it to a UMLS concept. Two possible reasons can be conjectured: (1) the larger the term the more likely the lexical variations among that term and a semantically equivalent term created by someone else; and (2) the more complex the term, the less likely it is to be in UMLS. It was our assessment that the complexity of the multi-word terms was necessary to preserve the semantics.

The development of standard terms for error types and causative factors using compositional knowledge representation approaches can be facilitated by an understanding of the structure of the terms. The results from the partial matching algorithm can help in identifying the semantic types of constituent concepts and their relationships for various term categories.

References

1. Kohn LT, Corrigan JM, Donaldson MS, editors. To err is human: building a safer health system. Washington, D.C: National Academy Press; 2000.
2. Kaplan HS, Callum JL, Rabin Fastman B, et al. The Medical Event Reporting System for Transfusion Medicine: will it help get the right blood to the right patient? Transfus Med Rev 2002;16(2):86-102.
3. Staender S, Davies J, Helmreich B, et al. The anaesthesia critical incident reporting system: an experience based database. Int J Med Inf 1997;47(1-2):87-90.
4. Bagian JP, Gosbee J, Lee CZ, et al. The Veterans Affairs root cause analysis system in action. Jt Comm J Qual Improv 2002;28(10):531-45.
5. Studdert DM, Thomas EJ, Burstin HR, et al. Negligent care and malpractice claiming behavior in Utah and Colorado. Med Care 2000;38(3):250-60.
6. Leape LL. Reporting of adverse events. N Engl J Med 2002;347(20):1633-8.
7. Stearns MQ, Price C, Spackman KA, et al. SNOMED clinical terms: overview of the development process and project status. Proc AMIA Symp 2001:662-6.
8. American Medical Association. CPT 2003 Prof. Ed. Chicago, IL: AMA Press; 2003.
9. International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM). 6th ed. Baltimore, MD: The Centers for Medicare & Medicaid Services; 2002.
10. Sangster W, Patrick T. Talking about medical errors: the void in existing controlled terminologies. Proc AMIA Symp 2002:1152.
11. Lindberg DA, Humphreys BL, McCray AT. The Unified Medical Language System. Methods Inf Med 1993;32(4):281-91.
12. National Coordinating Council for Medication Error Reporting and Prevention. Taxonomy of Medication Errors: United States Pharmacopeia; 1998. Available at http://www.nccmerp.org/taxo0731.pdf.
13. Zeng Q, Cimino JJ. Mapping medical vocabularies to the Unified Medical Language System. Proc AMIA Annu Fall Symp 1996:105-9.
14. McCray AT, Bodenreider O, Malley JD, et al. Evaluating UMLS strings for natural language processing. Proc AMIA Symp 2001:448-52.